

Development of Classical Tamil Digital Library: CIIL Experience

**B.A.Sharada Ph.D.,
Librarian
Central Institute of Indian Languages
Manasagangotri, Mysore-570 006, INDIA
sharada@ciil.stpmyssoft.net**

**Manju Naika
Classical Tamil Digital Library
Central Institute of Indian Languages
Manasagangotri, Mysore-570 006, INDIA
Manjunayak1977@yahoo.co.in**

Abstract

The present study describes the establishment and develops a digital library in Classical Tamil. An overview of the concept “Classical Tamil” is explained. The need for developing classification schedule in Classical Tamil was much felt and a module is presented. The study also proposes the creation of bilingual metadata with sufficient search facility in the original language using MARC format and Unicode. It will also discuss the digitising process, storage, retrieval, dissemination, preservation, etc in original language

1. Introduction

In the present digital era, bringing the knowledge database in Indian languages in the internet is a challenging factor. However, keeping in view, collection of English documents, setting up a digital library is relatively an easy task using the current data base technology and availability of qualified library professionals, Standard Cataloguing Format, various components and tools available today. But, it is not the same for Indian languages.

The present study attempts to describe the efforts of developing a Classical Tamil Digital Library (**CTDL**) having the collection on Ancient Tamil Studies. Following are some of the aspects involved in building this library:

- Developing collection in Ancient Tamil (Print/electronic)
- Developing classification schedule in Classical Tamil
- Online Classification
- Cataloging of these documents by using MARC Standard Format and Unicode
- To compile an exhaustive catalog of works on ancient Tamil studies available as manuscripts, print version, audio, video, multimedia formats found in India and overseas libraries.
- To prepare subject wise annotated bibliographies (for grammar, literature, folklore etc)
- Digitization of Ancient Tamil Works

2. Classical Tamil Digital Library Background:

Language occupies an important position in the history of Indian cultural traditions. The concept Classical status to a language is not new. In the Ministry of Human Resource Development (MHRD), Government of India, the languages that have already obtained the Classical status are, Arabic, Prakrit, Pali, Persian and Sanskrit. The Central Government declared Tamil as one of the classical languages on September 17, 2004. Over the millennia, Tamil has contributed to the making of Indian cultural heritage. The meeting held in the MHRD entrusted the Central Institute of Indian Languages (CIIL), Mysore an apex institution under the Language Division, MHRD with the

responsibility of implementing, monitoring and evaluation of the scheme.

The first national consultative meeting of the Central Scheme for Classical Tamil was held on Nov 18-19th 2005 and the Central plan Scheme for Classical Tamil, Tamil Language Promotion Board was held on 10.12.2005 at Central Leather Research Institute, Chennai. In order to implement the scheme, the ten member committee put forth their suggestions. One of such suggestions was to set up “A National Resource Library in Classical Tamil” (both print and digital). Based on this, the First Meeting of the Project Advisory Committee (PAC) on Digital Library for Ancient Tamil Studies was held on 1 & 2 February 2006, at CIIL, MYSORE. The Project Advisory Committee consisted of the following members who were experts in computer science, virtual library, Tamil language, library and information science:

Coordinator: Dr. B.A.Sharada

1. Dr Anto Peter, Chennai
2. Prof. V. Jayadevan, Chennai
3. Dr. Ponnaivaikko, Chennai
4. Dr S. Parthasarathy, Chennai
5. Dr A. Neelameghan, Bangalore

“Digital Library for Ancient Tamil Studies” is getting developed on proposals rendered by the PAC members.

2.1 Project Requirements

The already available VTLS’s Virtua software in the CIIL library which is compatible with all Indian languages with UNICODE and MARC 21 support is being used for CTDL also. VTLS stands for Visionary Technology in Library Solutions, from Virginia USA. The IT infrastructure of CIIL library includes:

- Solaris Sun Fire-280 server and 1 Intranet server
- 2 Five KV UPS, 40 Systems, 1 CD Share device, 1 Secondary storage device
- 2 Scanners (HP Scanjet 8290 and Minolta PS 7000) and 2 Barcode readers.
- CIIL has LAN with 350 nodes as well as V-SAT connection.

3. Developing collection in Ancient Tamil

The CTDL work started in the month of February 2006 itself by collecting the catalogues from different publishers in Tamil Nadu. The Tamil language experts were consulted to select the titles that come under Ancient classical Tamil field and those books were acquired to the library. Some scholars personally visited different publishers and directly purchased the relevant books.

4. Classification of Ancient Tamil documents:

The critical edition of ancient Tamil works will include 41 works namely

- 1) Tolkappiyam
- 2) Padenen Mel kanakku (18 Upper Classics)
- 3) Padenen Kizhk kanakku (18 Lower Classics)
- 4) Silappadiharam
- 5) Manimegalai
- 6) Muutollayiram
- 7) Iraiyanar kalaviya.

The chronological order of the 41 works is essential for proper organization. For shelving the books on the racks, same chronological order has to be followed. For this purpose, the existing modified classification system was not suitable. Hence a new classification scheme for CT was developed based on DDC. Since most of the Ancient Tamil works come under poetry and anthology, one Main class was derived as T 821.08, which has 41 categories and in each category it has 4 sub divisions i.e. Text, Commentary, Translation, Criticism. Methodology followed for the same is as follows:

4.1 Idea Plane

- Major groups in Ancient Tamil literature are identified.
- Identified groups are classified into various categories.
- Main topics are identified as headings of the categories
- Sub groups are identified in each of the major groups.
- Topics are hierarchically arranged along with their sub topics.
- Most of the ancient Tamil literature is rendered in the form of poetry.
- Number for poetry from modified DDC is chosen.

Preparation of the schedule involved discussions with Tamil scholars who were well versed in both modern Tamil as well as Ancient Tamil.

4.2 Verbal Plane

Ancient Tamil Texts are identified in 41 categories. They are arranged under major Headings. By examining the available literature books it is observed that each subgroup has similar sections such as:

1. Text
2. Commentary
3. Translation
4. Criticism

For example if 'ettu thokai' has eight separate anthologies under different heading, each anthology has the above four divisions. They are considered as common categories and introduced in each category with a mnemonic number.

After collecting the concepts, they are hierarchically arranged and the number worked out. The main headings are as follows:

- | | |
|------------------------|----------------------|
| 1. Tolkappiam | Grammar |
| 2. Ettuthokai | eight anthologies |
| 3. Pathu pattu | ten anthologies |
| 4. Pathinenkizhkanakku | eighteen anthologies |
| 5. Aimperunkappiyam | big epics |
| 6. Aimchirukappiyam | epics |

4.3 Notational Plane

The first one 'tolkappiyam' is a grammar. The class number is available in the modified classification schedule for linguistics¹ –

T 422.13 where T stands for Tamil.

420 Dravidian Linguistics

422 Branches and Schools

422.1 Descriptive School

422.13 Grammar

In DDC, 400 relates to Language, 410 to Linguistics and 800 to Literature. Wide gap of 400 numbers is there in between. The problem exists with Literature related topics. One plus point is that in ancient

Tamil, literature has been rendered in poetry form, most of them being anthologies. So, the main number that is to be worked out in the notational plane is the class number 'T821.08'.

Where, T = Tamil

820= Literature

821= Poetry

821.08 Collection of Poetry

Keeping in mind the numerical representation in anthologies, an attempt was made to put it in the following form:

Ettam thokai (eight anthologies) T821.08 8

Pathu paattu (ten anthologies) T821.08 10

However, when books are arranged on the shelves, item 6 comes first and chronology is not maintained. Hence the following method was used for notation.

T422.13 Tolkappiam

T422.131 Text

T422.132 Commentary

T422.133 Translation

T422.134 Criticism

T821.081 Classical Tamil Literature

T821.081 1 Ettuthokai

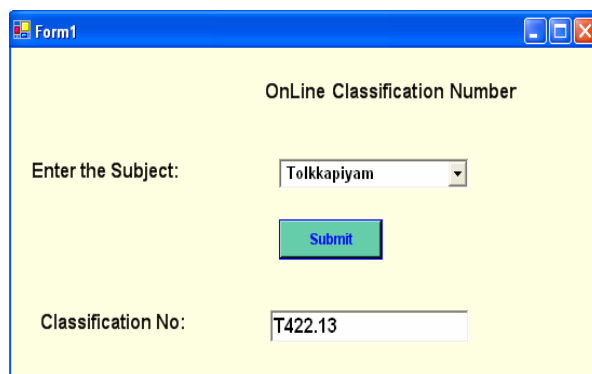
All the eight subclasses are given separate number in the hierarchy. The common category numbers are common to subclasses. A sample version of the schedule is available in the site <http://www.ciillibrary.org/ct/index.html>. A tag in the MARC format to enter the local number is created in the VTLS worksheet under 900 tag.

4.4 Online Classification

The classification schedule prepared for Ancient Tamil has been keyed in using UNICODE and a programme developed to develop database and retrieve the classification number, given the topic. To

make it user friendly, user interface with search button has been introduced. The online classification schema is given below:

Schema 4



Form1

OnLine Classification Number

Enter the Subject: Tolkkapiyam

Submit

Classification No: T422.13

5. Cataloging:

Bibliographical metadata was entered in MARC21 format as a follow up action to accessioning work by using Virtua software. While cataloging this Tamil collection, both Tamil script and Roman transliteration has been used to the following elements i.e. personal name, corporate name, meeting name, title and statement of responsibility, Place of publication, Publisher name and Subject heading. Machine transliteration to roman script is in progress. The newly developed classification schedule enabled to give proper subject heading in both English and Tamil. Bilingual metadata is created in both Tamil script and Roman transliteration.

6. Information Retrieval in Tamil Language:

Information can be retrieved through OPAC. There are various types of searches as Basic, Key word and advanced search. Since the language of the document is also entered while cataloguing, language wise search is also possible.

As in English language, the any type of searches could be done in Tamil also by entering in Tamil itself. From Search Menu we can search various types of searches such as Author, Title, Subject,

Language, etc. In the browse tab of the search window, if any type of search is selected as search category, and in the text box if the search string is entered in Tamil, it will display search result in Tamil Language itself. For example:

Author

பாரதிதாசன் (Bharathidasan)

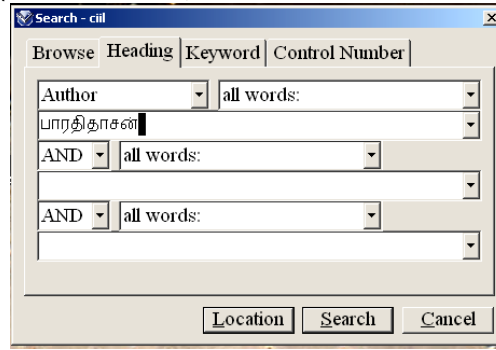


Figure-1: Search Window

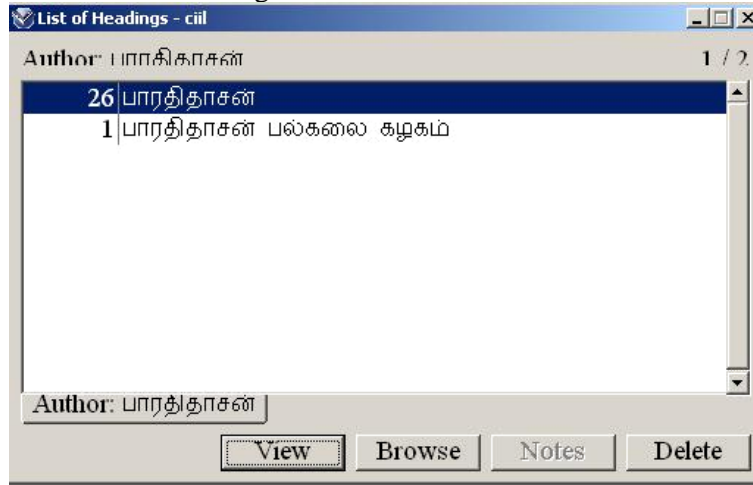
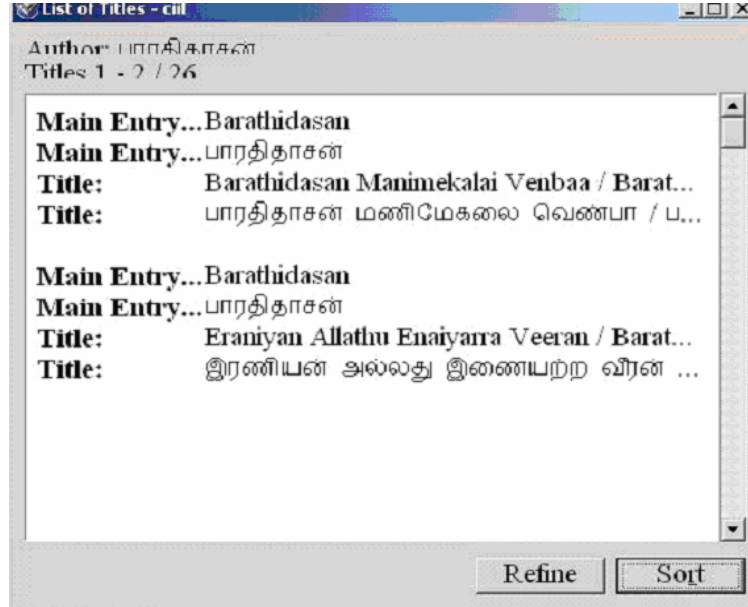
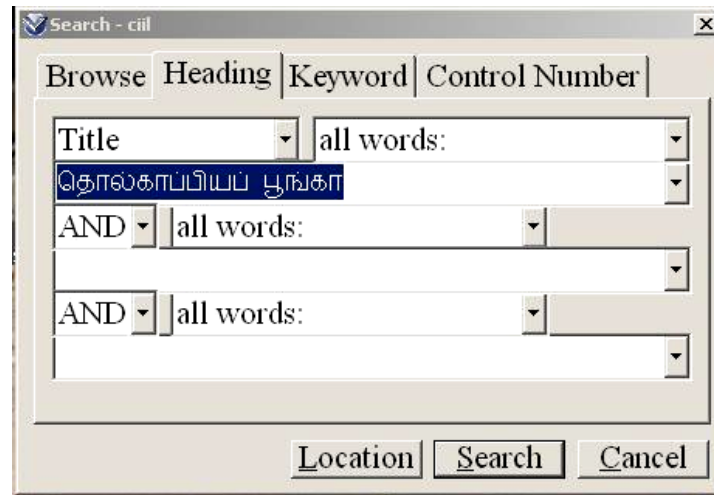


Figure 2: Retrieval Window 1



Title Search

தொல்காப்பியப் பூங்கா (Tholkappiya Poonga)



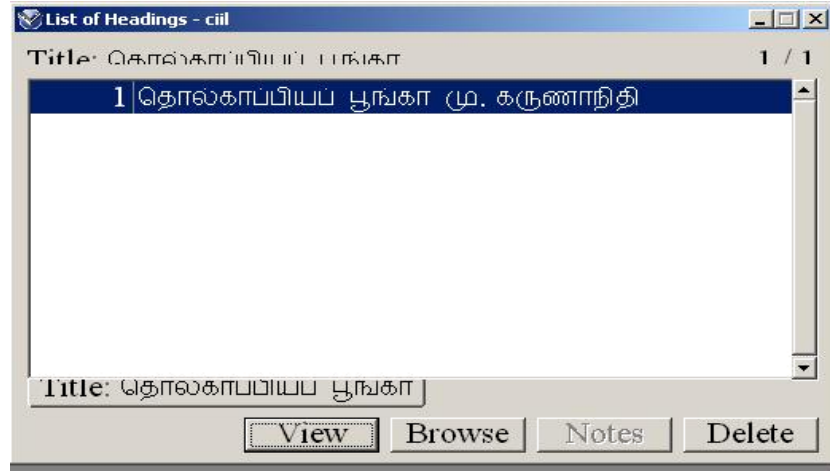


Figure 5: Title Retrieval Window

2.5.1 Subject Search

திருக்குறள் (Thirukkural)

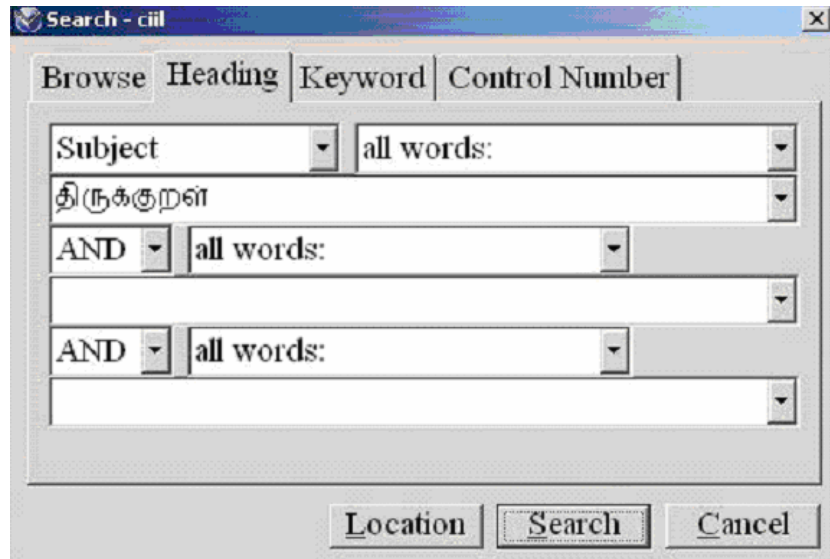


Figure 6: Subject Search Window

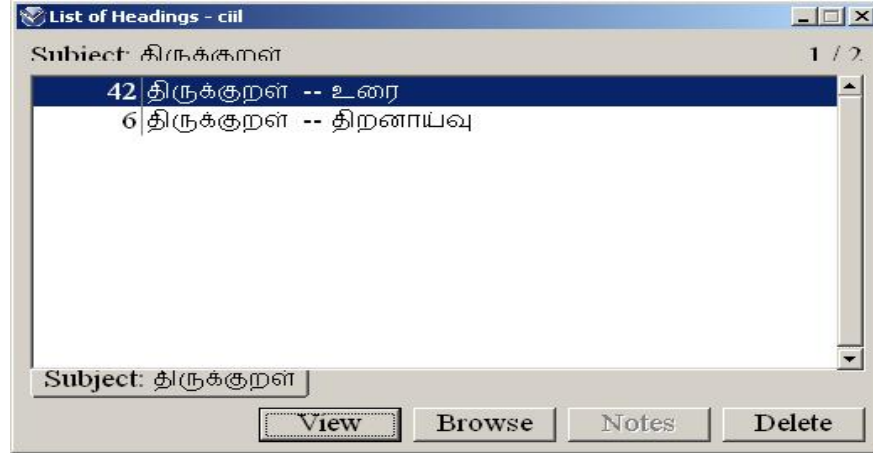


Figure 7: Subject Retrieval Window 1

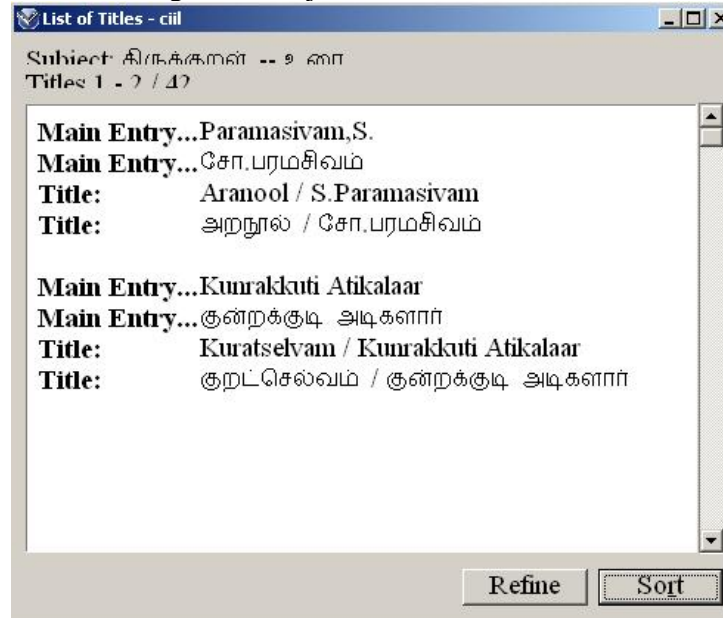


Figure 8: Subject Retrieval Window 2

7. Digitization:

To feed ancient Tamil literature into the internet for example: To produce electronic texts of *Tolkāppiyam*, Sangam literature, *Tirukkural*, *Silappathikāram*, *Kambarāmāyanam* etc. on the model of the electronic Bible text

Before starting digitization work, some of the issues were discussed with Tamil scholars to know what to digitize, which material are required for Research and Education purpose in this field. Based on the inputs from Tamil scholars, it was decided to digitize material without violating copyright act. Since classical works come under ancient period (6th Century AD), copyright has expired for majority of this collection. Wherever necessary, we have to take permission from the respective copyright owners. Minolta PS7000 Scanner and Abby Fine reader OCR 8.0 software is being used for this purpose.

Also, more than two hundred documents which were available freely on the web were copied to our database and a link to the full text was given to the catalogue.

The above projects will enable us to establish a National Reference Library for Tamilology as to make available all the pertinent works in one place; to digitize all those works and put it in the internet so as to be accessed by any one from anywhere in addition to the available ones.

Conclusion

It is through Virtua Integrated library management software along with international MARC format and Unicode support, information management in Tamil could be materialized successfully by providing a perfect module to create a digital library. While developing the Classical Tamil Digital library, any one can utilize the digital content for various academic and research purposes. The present study analyses and evaluates every step in the process of development of Digital library from acquisition, storage, retrieval and dissemination of information connected to ancient Tamil collections.

Future

Explore or adopt if already done, the powerful information retrieval procedure to the digitized documents.

References

1. Biligiri, H.S and Sulochna, C.R. (1973). *Classification for Linguistics and Languages: A modified Approach on Decimal Classification* (pp.31). Mysore: Central Institute of Indian Languages.
2. Kulandai Swamy, V.C. (2005). *Tamil among the Classical Languages of the World*. Chennai: Pavai Publications.